

LAR VIDEO : CODAGE VIDEO SANS PERTE A SCALABILITE SEMANTIQUE

Samir AMIR¹

Erwan FLÉCHER¹

Marie BABEL²

Olivier DÉFORGES²

UMR CNRS 6164 IETR Groupe Image et Télédétection

INSA de Rennes

20, avenue des buttes de Coësmes, 35043 Rennes, FRANCE

¹{samir, eflecher}@ens.insa-rennes.fr

²{mbabel, odeforge}@insa-rennes.fr

Concours Jeune Chercheur : Oui

Résumé

Baptisé “LAR Video”, le schéma proposé dans cet article décrit un nouvel algorithme de compression vidéo sans perte avec scalabilité sémantique. Propre au codage vidéo, une étape d’estimation de mouvement est effectuée afin de produire une image d’erreur résiduelle. L’idée ici est d’appliquer sur cette erreur, une décomposition pyramidale issue d’une méthode de codage scalable d’images fixes, le LAR-APP. Résultantes d’une prédiction inter/intra niveau, les erreurs d’évaluation sont transmises progressivement. Le décodeur peut ainsi reconstruire les images de la séquence de façon scalable par niveau de résolution spatiale. Enfin au vu des résultats obtenus, nous pouvons affirmer que le schéma proposé offre en plus de la scalabilité, des performances intéressantes en compression.

Mots clefs

Partitionnement quadtree, scalabilité sémantique, codage vidéo sans perte multirésolution.

1 Introduction

Le codage vidéo est utilisé dans de nombreuses applications qui nécessitent le développement d’outils de codage efficaces et rapides afin de compresser au maximum le flux vidéo tout en conservant une qualité visuelle optimale. Ces applications doivent pouvoir faire face aux caractéristiques hétérogènes et variantes dans le temps des réseaux de transmission. Pour résoudre ce problème, il est intéressant de disposer d’un flux vidéo scalable [1]. La scalabilité désigne l’aptitude d’un algorithme de compression à représenter une source hiérarchiquement sur plusieurs couches complémentaires [2]. Tout d’abord, une couche de base permet la reconstruction des informations émises avec une qualité minimum. D’autres niveaux de raffinement se succèdent ensuite. L’application des traitements non réversibles pour le codage vidéo avec perte est nécessaire à l’obtention des forts taux de compres-

sion. En contre partie, ils peuvent détruire une partie essentielle de l’information ou faire apparaître des artefacts conduisant à une interprétation erronée de l’image. Ces dégradations ne sont pas tolérables pour certaines applications dédiées notamment la télé médecine et la production cinématographique. Par exemple, une échographie en l’absence de compression, nécessite un temps de transfert inconcevable sur les réseaux de communication classiques. A ce jour, rares sont les solutions qui mutualisent à la fois compression sans perte efficace et option de scalabilité. Inspiré par cette constatation, un nouveau schéma d’un codage vidéo scalable basé sur une approche pyramidale de compression d’image fixe (LAR-APP) a été élaboré.

La méthode LAR (Locally Adaptive Resolution) fondée sur une représentation de l’image à taille de bloc variable définit une technique efficace de compression sans perte [3]. Son principe de codage a inspiré trois approches pyramidales de compression d’images fixes, le LAR-APP, l’Interleaved S+P et le RWHT+P [4, 5, 6]. Après avoir évoqué les particularités de ces trois méthodes (§2), nous décrivons le principe et les performances d’un schéma de codage vidéo sans perte qui se base sur le codage scalable de l’erreur de prédiction par le LAR-APP (§3). Enfin les résultats obtenus par le même schéma mais dans le cas de la non réversibilité sont présentés dans paragraphe 4.

2 Méthode LAR sans perte pour image fixe

2.1 Schéma général

Considérant l’image à compresser comme la superposition d’une information globale et de la texture locale, il est alors possible de décider, à partir d’un même schéma d’encodage, de fournir une image compressée à bas débit (information globale), ou de l’enrichir en texture (ajout de l’image d’erreur). Le principe de l’encodage en deux couches est la base de la méthode LAR et offre naturellement au moins deux niveaux de progressivité (Fig. 1).

La première étape du codage LAR d’une image consiste

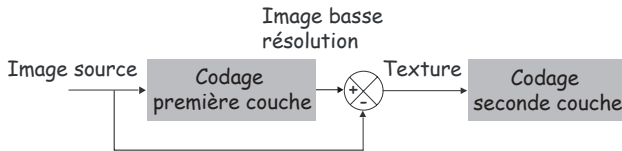


Figure 1 – Schéma général du codeur LAR : deux couches de codage

en l'obtention d'une image basse résolution (en termes de qualité visuelle). Cette phase est subordonnée à la création d'un partitionnement quadtree de l'image originale où la taille de chaque bloc est définie après évaluation de l'activité locale. Il en résulte une image de blocs de taille variable affectés de la luminance moyenne de son contenu. Ce pavage offre à la fois un lissage des zones homogènes mais aussi un bon rendu visuel des contours. Par la suite, nous définirons la partition quadtree par $QP^{[N_{max}..N_{min}]}$, où N_{max} et N_{min} représentent la taille maximale et minimale autorisée des blocs carrés. En l'absence de seconde couche, ce codeur vise clairement les forts taux de compression. L'image reconstruite bas débit est visuellement acceptable en raison de l'adaptation au contenu, du support de l'image des blocs par le quadtree. Pour des débits plus élevés, la première phase est suivie d'une couche de raffinement qui, si aucune quantification n'est réalisée, permet une compression sans perte de l'information. Ainsi, la propriété de codage par scalabilité sémantique est rendue possible par la mise en oeuvre conjointe d'une solution à deux couches (progressivité par palier) et du partitionnement quadtree. Soucieux d'enrichir ce principe de codage, trois approches scalables reposant sur la méthode LAR ont ensuite vu le jour.

2.2 Approches pyramidales

Le LAR-APP [4], l'Interleaved S+P [5] et le RWHT+P [6] sont trois méthodes de codage d'images fixes développées au sein du laboratoire. Il s'agit d'algorithmes unifiés de compression avec ou sans perte combinant prédiction dans un contexte enrichi et scalabilité par niveau de résolutions et de détails. Ces trois méthodes reprennent la décomposition en deux couches du LAR et y apportent en plus la multirésolution. Une décomposition pyramidale est effectuée selon deux descentes successives, contraintes par une partition quadtree préalablement calculée. La première descente opère un raffinement uniquement sur les contours et modélise les zones homogènes par de grands blocs. Le principe de la division conditionnelle est illustré par la figure 2 pour des tailles de blocs allant de 2×2 à 8×8 . Prenant en considération le partitionnement régulier quadtree du LAR, cette première décomposition pyramidale procure une scalabilité sémantique. A un niveau donné, la résolution adaptée au contexte de l'image accroît la qualité (sur critère visuelle) de l'imagette reconstruite et renforce ainsi l'efficacité du codage scalable de l'image des blocs. Si la première passe réalise la reconstruction de l'image

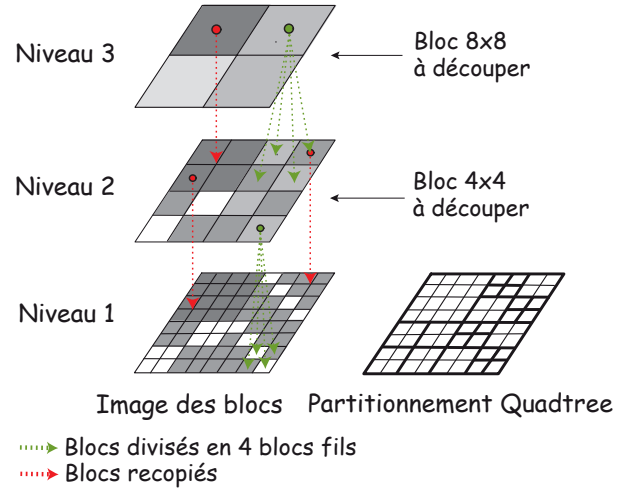


Figure 2 – Décomposition de la pyramide de l'image des blocs conditionnellement au pavage quadtree $QP^{[8..2]}$

basse résolution (LAR) par progression sémantique, la seconde permet de récupérer la texture locale. Ainsi à chaque niveau de la pyramide, les blocs ignorés par la première descente sont décomposés par la couche de raffinement.

La modélisation de contexte est devenue un facteur clé d'efficacité des algorithmes de compression sans perte [7]. En effet, l'entropie brute d'une image peut être réduite lorsque des classes de symboles propre, à différentes loi de probabilité, peuvent être isolées. L'observation du comportement de nos algorithmes vis-à-vis de l'entropie engendrée a permis la séparation des lois de probabilité en deux niveaux : la taille de blocs et le niveau de décomposition dans chacune des deux passes. De ce fait, grâce à la nature même du schéma de codage, une modélisation de contexte implicite est réalisée.

A ce stade, il est raisonnable d'estimer que nos trois approches pyramidales reposent sur une même base algorithmique opérant selon deux passes successives. Si la première décrit une pyramide d'imagettes LAR dont la qualité évolue conformément au principe de scalabilité sémantique, la seconde offre un enrichissement progressif en texture. Bénéfique d'un point de vue entropique, la modélisation de contexte implicite est un autre aspect largement exploité par ces trois méthodes.

2.3 Méthodes de décomposition

Après avoir brièvement présenté les similitudes de ces trois approches, il est maintenant intéressant d'en évaluer les disparités basées sur le processus de décomposition d'un bloc en quatre fils dans le contexte d'un codage réversible. La prédiction MICD scalable du LAR-APP repose pour sa part sur une succession de passes visant à tirer profit des informations contextuelles présentes aussi bien dans le domaine causale (prédiction intra-niveau) que dans l'image sous-échantillonnée (prédiction inter-niveau). Contrairement au LAR-APP, l'Interleaved S+P et le RWHT+P ne se

restreignent pas au domaine spatial mais oeuvrent dans des domaines transformés plus appropriés à l'obtention de taux de compression élevés. Dans son principe, l'Interleaved S+P est associé à une implantation particulière de la transformée en S. Plus précisément, son espace de prédiction résulte de l'application de la transformée en S 1D sur deux vecteurs constitués chacun des deux pixels diagonalement adjacent d'un bloc 2×2 donné. L'originalité de cet algorithme tient dans le fait qu'il est possible de prédire les coefficients de la transformée au moyen de deux pyramides entrelacées. Pour sa part, la méthode RWHT+P met en oeuvre une version réversible de la transformée Walsh Hadamard appliquée à des blocs 2×2 .

L'Interleaved S+P et le RWHT+P sont deux méthodes pyramidales extrêmement efficaces de compression sans perte, surpassant largement le LAR-APP et les méthodes de l'état de l'art telles que CALIC [8] et S+P [7]. Malgré cela, c'est le LAR-APP qui a été implanté dans un premier temps dans notre codeur vidéo sans perte en raison de sa simplicité de mise en oeuvre. L'une de nos perspectives proches vise à le substituer par l'Interleaved S+P et le RWHT+P afin d'accroître les performances de codage.

3 Principe du codage vidéo LAR sans perte

L'estimation et la compensation de mouvement sont deux étapes majeures des standards de codage vidéo. En exploitant la redondance temporelle entre images successives, un fort taux de compression peut ainsi être obtenu. L'idée de base du codeur vidéo est de prédire l'image courante par une référence compensée en mouvement. La première étape consiste à diviser l'image en blocs réguliers (MPEG-4) ou selon un découpage irrégulier (H.264 [9]). Au moyen d'un critère de distortion et pour chaque bloc, l'estimateur de mouvement recherche le meilleur appariement dans une image de référence. La distance entre les deux blocs est représentée par un vecteur de mouvement qui est associé au bloc courant et envoyé au décodeur. A ce stade, l'erreur résiduelle résultante de la différence entre l'image compensée et l'image à coder doit être efficacement transmise. Pour y parvenir, une transformée est appliquée sur l'erreur afin de profiter de la redondance spatiale présente. Classiquement les moyens utilisés par les standards permettent une forte compression mais n'offrent pas de solution performante au codage sans perte.

La construction d'un schéma de codage vidéo par l'approche pyramidale prédictive (§2) vise donc à répondre à deux attentes fondamentales. La première consiste à proposer une méthode nouvelle scalable par niveau en résolution et en qualité. Le second objectif est de proposer à l'utilisateur une solution simple et unifiée de compression avec ou sans perte. Le LAR-APP appliqué à l'erreur résiduelle en est une réponse : un seul algorithme est utilisé à des fins de codage réversible ou non (Fig. 3).

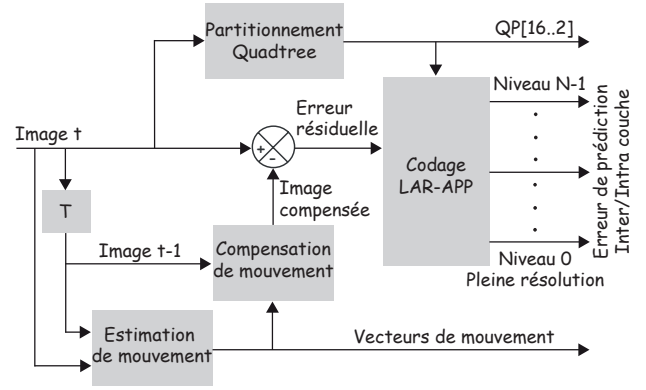


Figure 3 – Schéma général du codeur vidéo LAR sans perte avec compression de l'erreur résiduelle sur N niveaux par le LAR-APP

3.1 Estimation de mouvement

La mise en correspondance de blocs est une solution simple et efficace d'estimation de mouvement. Ce procédé approxime le mouvement apparent entre deux images par un modèle translationnel propre à chaque bloc issu du découpage de l'image courante. Une recherche exhaustive du meilleur appariement donne la solution optimale (au sens d'un critère de ressemblance) mais s'avère demande d'une grande puissance de calcul. De la famille des techniques de recherche par zone, l'estimateur de mouvement EPZS (Enhanced Predictive Zonal Search) [10] surmonte ce problème en restreignant l'espace de recherche à partir d'une connaissance à priori du déplacement. La prédiction de déplacement d'un bloc donné est issue du mouvement évalué dans son voisinage causal mais aussi dans une image de référence. Même si le codage de l'erreur résiduelle n'a pas encore été détaillée, il peut être intéressant d'analyser l'influence de la taille des blocs sur l'entropie de cette erreur. Le graphe ci-dessous (Fig. 4) illustre l'ascendance d'une grille régulière de blocs 16×16 , 8×8 et 4×4 sur le codage de l'erreur par l'approche pyramidale LAR-APP. Les entropies mentionnées résultent du codage sans perte des images de la séquence Foreman.

La diminution de l'entropie de l'erreur engendrée par une prédiction plus fine, s'accompagne naturellement d'une augmentation du coût des vecteurs de mouvement. Dans un contexte de codage à bas et moyen débit, des solutions [11] comparables à celle adoptée par H.264 [9] s'avèrent appropriées. Elles reposent généralement sur un partitionnement optimal de l'image au sens du mouvement en considérant simultanément le coût des vecteurs de mouvement et celui de l'erreur résiduelle. Or appliqué au codage réversible, il est évident que le gain apporté par ces solutions serait dilué dans la dynamique entropique de l'erreur. La solution retenue dans notre codeur est d'estimer le mouvement sur un grille régulière de bloc 8×8 (meilleure solution dans notre exemple).

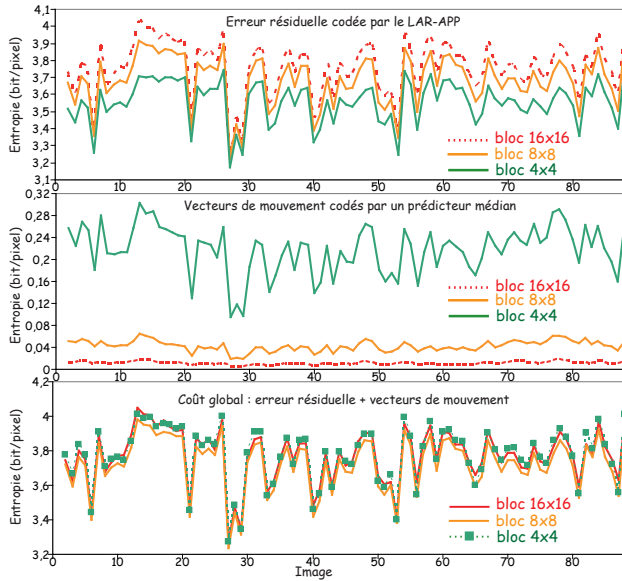


Figure 4 – Influence de la partition régulière associée au modèle de mouvement sur l'erreur résiduelle codée par le LAR-APP sans perte (séquence Foreman)

3.2 Codage de l'erreur résiduelle

Afin de tirer profit de la scalabilité sémantique et de la modélisation de contexte, la décomposition (sur critère spatial) est ici appliquée sur des images directement extraites de la séquence. Le partitionnement de l'image ainsi que la décomposition en flux scalable de l'erreur résiduelle par le LAR-APP peuvent être modélisés par un schéma bloc (Fig. 5). Conformément à la méthode décrite dans le paragraphe 2, la compression sans perte de l'erreur résiduelle consiste à établir un codage pyramidal total. Chaque bloc du pavage LAR est décomposé jusqu'à l'obtention de la pleine résolution, selon le principe de décomposition pyramidal et de prédiction dans un contexte enrichi. Le codage par scalabilité sémantique impose que l'information liée aux petits blocs soit dissociée de celle issue du reste de l'image, de telle sorte que la première passe décrive exactement les contours forts. Ainsi le schéma général définit un niveau de décomposition minimal relatif à la taille des blocs. Associé à une partition $QP[16..2]$, le schéma (Fig. 5) met en lumière ce principe dans le cas où pour un niveau n de la pyramide, la taille minimale de décomposition autorisée est 2^n . Par conséquent, le niveau 0 de la pyramide est entièrement traité lors de la seconde descente. En raison de la similarité du processus d'évaluation au niveau du récepteur, seules les erreurs de prédictions inter/intra couches transitent conformément au principe de l'emboîtement de flux.

Au décodeur, le processus de reconstruction scalable des images opère en deux temps (Fig. 6). A un niveau donné de la pyramide, l'erreur résiduelle est tout d'abord décodée puis sommée à une image compensée de même dimension. Pour cela une pyramide d'images compensées est

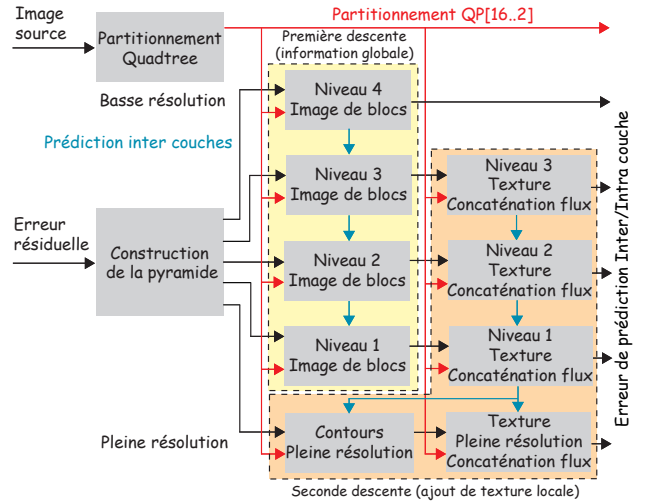


Figure 5 – Codage scalable de l'erreur résiduelle par l'approche pyramidale prédictive LAR-APP

élaborée sur le même principe que celle de l'erreur résiduelle. Naturellement, le décodage par couche scalable nécessite une étape de prédiction/compensation de mouvement progressive. Même si pour l'instant cette phase n'est pas encore réalisée limitant ainsi l'entièrement progressivité de notre schéma de codage, des applications peuvent lui être destinées. Dans le cadre d'une compression réversible, la pyramide se doit d'être complète afin de retrouver après l'opération de décodage, une image originale compressée sans perte. Ainsi l'utilisation des images décodées comme moyen de prévisualisation des images de la séquence à différents niveaux de résolution est une des fonctionnalités offerte par notre décodeur.

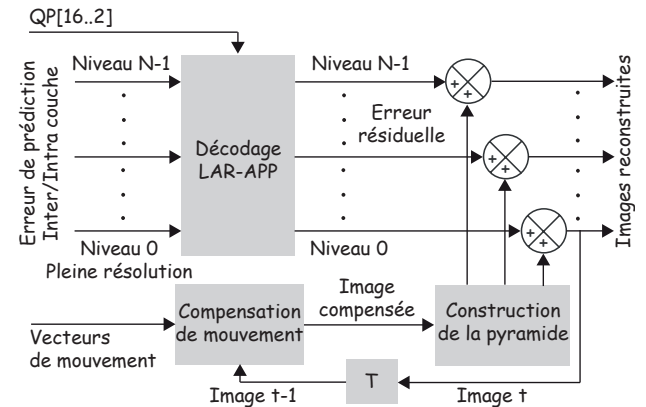


Figure 6 – Schéma général du décodeur vidéo LAR sans perte avec décompression de l'erreur résiduelle sur N niveaux par le LAR-APP

Les figures 7 et 8 présentent les images successivement obtenues lors du décodage d'une image de la séquence Football par l'approche pyramidale prédictive. Elles décrivent respectivement la pyramide de l'erreur

résiduelle, ainsi que les images reconstruites progressivement selon le processus précédemment décrit.

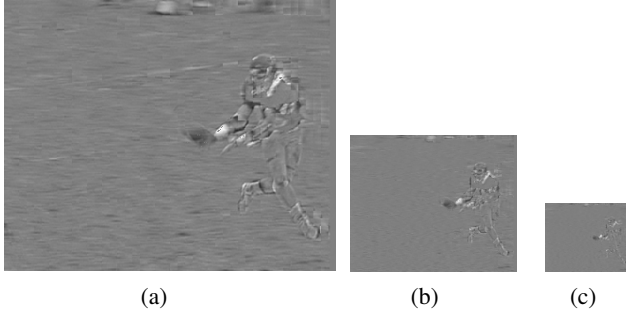


Figure 7 – Images d’erreur intermédiaires obtenues lors de la décomposition pyramidale. (a) Pleine résolution, sans perte à 3.80 bpp (b) niveau 1 à 1.01 bpp (c) niveau 2 à 0.37 bpp

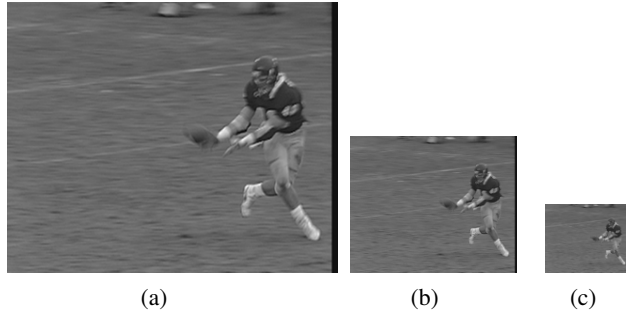


Figure 8 – Reconstruction progressive de l’image source jusqu’à la pleine résolution à partir de la pyramide de l’erreur résiduelle. (a) Pleine résolution, sans perte (b) niveau 1 (c) niveau 2

La figure 9 montre la reconstruction d’une image pleine résolution à des niveaux intermédiaires en qualité. Comme expliqué précédemment, elle résulte de l’addition de l’image compensée pleine échelle et dans le cas présent d’une erreur intermédiaire. Dans le premier cas, l’erreur est obtenue par interpolation de l’image issue de la première passe (image LAR du niveau 1). Dans le second cas, la pleine résolution de l’erreur est atteinte en décomposant uniquement les contours (blocs 2×2).

Après avoir présenté notre codeur vidéo ainsi que certaines images intermédiaires, il est maintenant intéressant d’en évaluer l’efficacité. Le tableau 1 synthétise les résultats obtenus lors du codage de diverses séquences tests par le LAR-APP sans perte. Les performances de l’approche pyramidale sur des images directement extraites de la séquence (APP Intra) et celles résultantes de l’erreur de compensation de mouvement (APP Inter) y sont comparées selon un critère entropique. Au vu des résultats, il est aisé de conclure que le codage en mode Inter surpassent l’APP Intra. Les entropies obtenues ici laissent présager du gain réalisé par l’implantation future de l’Interleaved S+P et du

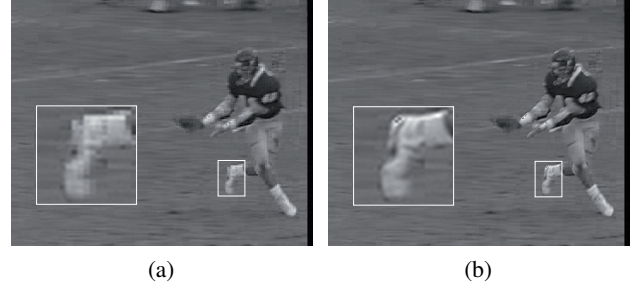


Figure 9 – Reconstruction intermédiaire au moyen de l’erreur issue de la première passe (a) erreur première passe + décomposition des contours (b) erreur première passe + interpolation

RWHT+P.

Séquence	Entropie (bpp) - Format CIF (352×288)			
	APP Intra	Erreur résiduelle	APP Inter	Vecteur
Foreman	4.30	3.66	3.69	0.042
Mother	3.60	2.91	2.85	0.035
Mobile	5.86	5.17	5.04	0.030
Football	4.09	4.33	3.81	0.081
Tempete	5.35	4.68	4.51	0.033
CoastGuard	5.36	4.66	4.40	0.025

Tableau 1 – Performances en mode Intra et Inter de l’approche LAR-APP sans perte sur différentes séquences tests

La figure 10 présente une description plus fine des résultats obtenus précédemment pour les séquences Football et MotherandDaughter. Elle illustre à la fois l’évolution temporelle de l’entropie brute de l’erreur résiduelle et celle engendrée par l’approche pyramidale sans perte. Tout en offrant une scalabilité par niveau, le LAR-APP permet une compression efficace.

4 Codage vidéo LAR avec pertes

De part sa nature non réversible, l’opération de quantification est naturellement bannie de notre schéma de codage sans perte. Par contre lorsque des dégradations sont tolérées, cet outil s’avère efficace afin d’obtenir des taux de compression importants. Classiquement les techniques de compression fondées sur l’optimisation débit/distorsion tentent de trouver le meilleur compromis entre coût de codage et détérioration de l’image d’un point de vue PSNR. Or des expérimentations ont montrées que le système visuel s’avère beaucoup moins sensible à des variations de luminance dans des zones de type frontière que dans des zones uniformes. Rappelons que notre partitionnement quadtree est piloté par l’activité locale, de grands blocs représentent des régions uniformes alors que des blocs de petits taille modélisent les contours. Ainsi une quantification conditionnée par la surface des blocs, va être associée au précédant schéma de codage à deux couches.

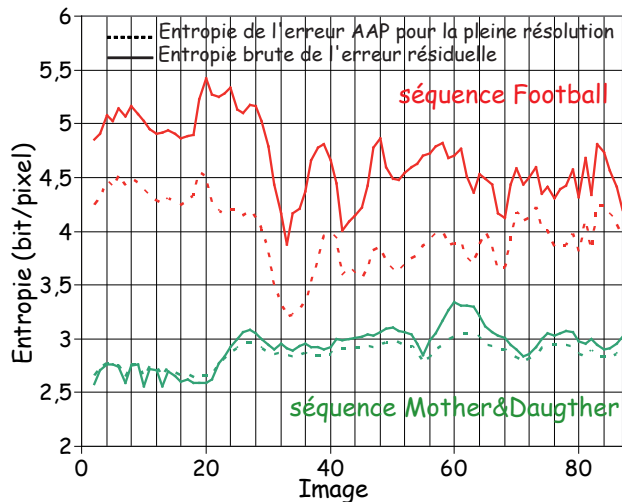


Figure 10 – Entropie résultante du codage par le LAR-APP sans perte comparée à l'entropie brute de l'erreur sur les séquences Football et MotherandDaughter

Pour des bas débits, le codage de l'erreur résiduelle par la première passe du LAR-APP associée à une quantification adaptative, offre une compression accrue tout en maîtrisant les dégradations engendrées. La figure 11 illustre ce principe sur une image pleine résolution (pr) de la séquence Football. Les pas de quantifications adoptés sont définis par $Q^T = [q_{16}, q_8, q_4, q_2, q_{pr}]^T = [2, 4, 8, 16, 32]^T$

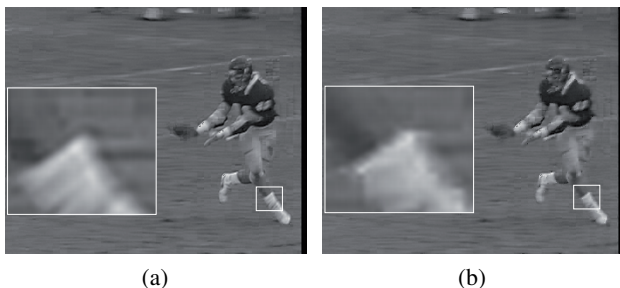


Figure 11 – (a) Image reconstruite à 37.12 dB par une erreur non quantifiée à 0.85 bpp. (b) Image reconstruite à 35.88 dB par une erreur quantifiée à 0.10 bpp

5 Conclusion

Le schéma de codage vidéo scalable présenté est une solution efficace de compression avec ou sans perte. Nous avons montré sur plusieurs exemples que les performances du schéma proposé, s'avèrent encourageantes pour la suite de nos travaux. L'objectif à court terme est de substituer le LAR-APP par l'Interleaved S+P et le RWHT+P, qui comme nous l'avons dit, reposent sur le même principe mais accroissent l'efficacité de compression. Appliquer un codage par une approche pyramidale sur l'erreur résiduelle n'est pas suffisant pour doter notre schéma d'une entière progressivité. La deuxième perspective est donc de rendre

la méthode intégralement scalable en incluant entre autre, un codage hiérarchique des vecteurs de mouvement.

Références

- [1] G. Marquant. *Représentation par maillage adaptatif déformable pour la manipulation et communication d'objets vidéo*. Thèse de doctorat, Université de Rennes 1, Rennes, Décembre 2000.
- [2] N. Cammas. *Codage vidéo scalable par maillages et ondelettes 2D*. Thèse de doctorat, Université de Rennes 1, Rennes, Novembre 2004.
- [3] O. Déforges et J. Ronsin. Region of Interest Coding for Low Bit-Rate Image Transmission. Dans *Proc. International Conference on Multimedia and Expo ICME'2000*, volume 1, pages 107 – 110, 30 July-2 Aug 2000.
- [4] M. Babel, O. Déforges, et J. Ronsin. Lossless and Lossy Minimal Redundancy Pyramidal Decomposition for Scalable Image Compression Technique. Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'03*, volume III, pages 249–252, Hong Kong, April 6-10 2003. Conference cancelled - Invited paper in ICME 2003.
- [5] M. Babel, O. Déforges, et J. Ronsin. Interleaved S+P Pyramidal Decomposition with Refined Prediction Model. Dans *IEEE International Conference on Image Processing, ICIP'05*, volume 2, pages 750–753, Genova, Italy, September 2005.
- [6] O. Déforges, M. Babel, et J. Motsch. The RWHT+P for an improved lossless multiresolution coding. Dans *European Signal Processing Conference, EU-SIPCO'06*, To be published, September 2006.
- [7] A. Said et W. A. Pearlman. An image multiresolution representation for lossless and lossy image compression. *IEEE Trans. on Image Processing*, 5 :1303–1310, September 1996.
- [8] X. Wu, N. Memon, et K. Sayood. A Context-based, Adaptive, Lossless/Nearly-Lossless Coding Scheme for Continuous-Tone Images (CALIC), 1995.
- [9] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, et T. Wedi. Video Coding with H.264/AVC : Tools, Performance, and Complexity. *IEEE Circuits and Systems Magazine*, 4(1) :7–28, First Quarter 2004.
- [10] A. M. Tourapis. Enhanced predictive zonal search for single and multiple frame motion estimation. Dans C.-C. J. Kuo, éditeur, *Proc. SPIE Vol. 4671*, p. 1069–1079, *Visual Communications and Image Processing 2002*, C.-C. Jay Kuo ; Ed., pages 1069–1079, Janvier 2002.
- [11] J. Zhang, M.O. Ahmad, et M.N.S. Swamy. Quad-tree Structured Region-Wise Motion Compensation for Video Compression. *IEEE Transactions On Circuits And Systems For Video Technology*, 9(5) :808–822, August 1999.